Warwick Economics Research Paper Series

Constructing Social Division to Support Cooperation

James P. Choy

February, 2016
Series Number: 1113
ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

This paper also appears as *CAGE working paper No: 266*

Constructing Social Division to Support Cooperation*

James P. Choy†

July 15, 2015

Abstract

Many societies are divided into multiple smaller groups. Certain kinds of interaction are more likely to take place within a group than across groups. I model a reputation effect that enforces these divisions. Agents who interact with members of different groups can support lower levels of cooperation with members of their own groups. A hierarchical relationship between groups appears endogenously in equilibrium. Group divisions appear without any external cause, and improvements in formal contracting institutions may cause group divisions to disappear. Qualitative evidence from the anthropological literature is consistent with several predictions of the model.

Keywords: Cooperation, Caste, Social Institution
JEL Classification Numbers: C73, O12, O17

*I am grateful to Mark Rosenzweig, Larry Samuelson, and Chris Udry for their advice and support throughout this project. Treb Allen, Priyanka Anand, David Berger, Gharad Bryan, Avinash Dixit, Tim Guinnane, Melanie Morten, Sharun Mukand, Motty Perry, Joe Vavra, and various seminar participants provided helpful comments. I acknowledge research funding from the Yale Economic Growth Center.

†Department of Economics, University of Warwick and CAGE. E-mail: j.choy@warwick.ac.uk

1 Introduction

Many societies are divided into multiple smaller groups. These divisions are especially salient in many developing countries, where the groups have names such as castes, tribes, or clans, but developed countries are divided as well, for example by race and religion. One stylized fact about group divisions is that people are more likely to interact in certain ways with members of their own groups than with members of different groups. Interactions that take place primarily within groups include trade (Greif 1993, Anderson 2011), mutual insurance (Grimard 1997, Munshi and Rosenzweig 2009, Mazzocco and Saini 2012), and job referrals (Munshi and Rosenzweig 2006). At first glance the lack of interaction between groups is puzzling, since the argument from the gains from trade suggests that people should seek to interact with the most diverse possible range of partners. In this paper, I argue that people may have a reputational reason to avoid interacting with members of different groups.

An example of social division due in part to reputation effects comes from Mayer's (1960) description of the caste system in the village of Ramkheri in central India. The central fact of the caste system, according to Mayer, is what he refers to as the commensal hierarchy, which prescribes who may eat with whom. There are five major caste groupings in the village, and members of higher ranked castes refuse to eat with or accept food from members of lower ranked castes, although members of lower ranking castes are willing to accept food from members of higher ranking castes. Mayer writes, "Eating the food cooked or served by a member of another caste denotes equality with it, or inferiority, and not to eat denotes equality or superiority." As eating together is one of the main ways to develop friendships, friendships are less likely to form across caste lines than within castes.

Whether people follow the rules of the hierarchy depends to some extent on whether other members of their caste can observe them. Mayer describes a member of an upper caste who was born in the village but who is working in the city of Indore. On a visit to the village, he is offered tea by a member of a lower caste, but he refuses, saying "I would willingly drink in Indore, but I must be careful not to offend anyone here." Similarly, Mayer describes a meal at a training camp for development workers held in the village, which is attended by delegates from many other villages. The delegates from other villages all eat together, while the delegates from Ramkheri sit separately in accordance with the caste rules. The Ramkheri delegates explain the situation, saying, "We could not sit with them here; but they, being away from their villages, were able to sit next to Muslims and even Harijans [members of the lowest Hindu caste]." According to Mayer this phenomenon is due to the greater difficulty in observing violations of caste rules that take place outside the village. Mayer writes, "The orthodox in Ramkheri know that the rules are being broken outside, [but] they are content not to investigate, so long as the matter is not given open recognition." Finally, after breaking the rules regarding caste contact, caste members are obliged to perform a ritual purification. However, whether the purification is in fact performed depends on whether the

violation is observed. Mayer writes, "Touching a Tanner [one of the lowest castes] is a more generally acknowledged matter for purification..., though it is admitted that many people would not do anything if they were not seen to touch." Thus people seem to follow the rules of hierarchy in part to preserve their reputations with members of their own castes.

Not all interactions between castes are penalized in Ramkheri. The Ramkheri caste system distinguishes between the sharing of different kinds of foods between castes. *Kacca* foods are foods cooked with water or salt. They include most daily staples. *Pakka* foods are foods cooked with butter. They are served at ceremonial occasions. The rules regarding *kacca* foods are much more stringent than the rules regarding *pakka* foods, and people are willing to accept *pakka* foods from members of lower castes from whom they would not be willing to accept *kacca* foods. My interpretation of this distinction is that sharing *kacca* food, which is eaten every day, is much more likely to lead to a deep, cooperative relationship than sharing *pakka* food, which is eaten only rarely.

To summarize, the Ramkheri caste system exhibits four important features. First, members of different castes do not interact in certain ways. Second, there is a hierarchy over castes, and members of higher ranking castes refuse to interact with members of lower ranking castes but not vice versa. Third, caste members follow the rules about non-interaction with other castes in part to preserve their reputations with members of their own castes. Fourth, the reputational penalties for interacting with members of other castes are more severe for those interactions which are most likely to lead to deep, cooperative relationships. I now outline a model that accounts for all of these features.

In the model, agents search over the community to find partners for cooperative relationships. If an agent cheats in any relationship, then the relationship breaks up and each partner to the relationship must search for a new partner. Search requires effort and hence is costly. Cooperation is maintained by the threat that any cheating agent will have to pay the cost of search, and the level of cooperation that any agent can support is inversely related to the search cost that the agent is expected to incur at the end of the relationship. Agents who expect to form matches with a larger fraction of potential partners pay lower search costs in expectation. Thus an agent who is expected to form matches with a larger proportion of the community can support a lower level of cooperation in any given relationship. Each agent is also a member of a payoff irrelevant group, and in equilibrium each agent interacts only with members of her own group. If an agent is observed to have formed a match with a member of a different group in the past, then it is believed that the agent will continue to accept matches both with members of her own group and with members of the other group in the future. Thus, agents who are observed to have interacted with members of different groups in the past are able to support lower levels of cooperation. This penalty for interacting with members of different groups is sufficient to prevent members of different groups from interacting in equilibrium. I refer to this state of affairs as group segregation. Group

segregation increases the level of cooperation that each agent can support compared to the situation without segregation, and if the benefits of cooperation are sufficiently important, then group segregation is welfare improving for the community as a whole.

The reputation mechanism yields novel theoretical insights. The first insight is that people may lose reputation with members of their own group by interacting with members of different groups. Specifically, people who interact with members of different groups are believed to be less trustworthy by members of their own group.

A second insight is that the reputation mechanism endogenously generates an asymmetry between different groups. Consider two groups, group 1 and group 2, and suppose that the reputation effect prevents members of group 1 from interacting with members of group 2. Members of group 1 do not interact with members of group 2 because it is believed that a member of group 1 who has interacted with a member of group 2 in the past will continue to interact with members of group 2 in the future. However, this belief is rational only if members of group 2 are willing to interact with members of group 1. Thus it must be the case that while members of group 1 are not willing to interact with members of group 2, members of group 2 are willing to interact with members of group 1. The groups are thus organized in a hierarchical structure, with higher ranking groups being unwilling to interact with lower ranking groups, but not vice-versa.

A third insight is that changes in formal contracting institutions could cause group segregation to break down. When deciding whether to accept a match with a member of a different group, a person must trade off the reputational penalty for accepting the match with the opportunity cost of rejecting the match in order to search for a relationship with a member of the same group. Improvements in formal contracting institutions increase the value of all relationships, even in the absence of the intertemporal incentives necessary to support cooperation. Thus, improvements in formal contracting institutions increase the opportunity cost of rejecting a match with a member of a different group and make it more likely that people will accept matches with members of different groups in spite of the reputational penalty. If formal contracting institutions improve sufficiently, group segregation is no longer an equilibrium.

In the literature the most closely related model to mine is Eeckhout (2006), which, like my model, features agents who are members of payoff-irrelevant groups, and who search over a community to find cooperative relationships. In Eeckhout's model, matched members of different groups do not cooperate at a high level even though relationships between members of different groups are potentially just as profitable as relationships between members of the same group. Intuitively, it seems implausible that people would consistently fail to realize the potential profits from their relationships in this way. I formalize this intuition by imposing a renegotiation-proofness concept called bilateral rationality, first introduced by Ghosh and Ray (1996). Bilateral rationality implies that if members of different groups do not interact,

then the potential profitability of relationships between members of different groups must be lower than the potential profitability of relationships between members of the same group. The reputation effect in my model lowers the potential profitability of relationships between members of different groups and ensures that my equilibrium is bilaterally rational.

Several other papers have provided reasons why relationships between members of different groups may be less profitable than relationships between members of the same group. Most of these papers hypothesize that some exogenous difference between members of different groups makes relationships between members of different groups less profitable. In the political science literature, divisions based on exogenous differences of this kind are referred to as "primordial" divisions, as discussed, for example, in Chandra et. al (2012). Two main kinds of primordial division have been described in the literature. The first kind of primordial division appears when members of different groups have different preferences. The simplest version of this idea is Becker's (1957) model of taste-based discrimination, in which people simply prefer to interact with members of their own groups. Other models with differing preferences between groups include Akerlof and Kranton (2000), in which people have preferences for expressing their identities by engaging in group specific behaviors, Bisin and Verdier (2000), in which people have preferences for passing on group specific traits to children, and Alesina, Baqir, and Easterly, in which members of different groups have different preferences over public goods. Tabellini (2008) constructs a continuous version of a model with differing preferences in which there is a metric over society and people have more altruistic preferences towards partners who are closer to them according to the metric. A second kind of primordial division appears if members of different groups have access to different communication technologies. The most obvious example of this kind of difference is if members of different groups speak different languages. Divisions generated by language differences are discussed by Lazear (1999) and Michalopoulos (2012). Even if members of different groups speak the same language, there may be more subtle differences in communication styles between groups that prevent members of different groups from communicating as effectively as members of the same group. This hypothesis is expressed most explicitly in Cornell and Welch (1996). Fearon and Laitin (1996), Miguel and Gugerty (2005), and Habyarimana et. al. (2007) argue that communication difficulties between groups inhibit information flows between groups and thus make contracting between members of different groups harder than contracting between members of the same group. Dixit (2003) constructs a continuous version of this model in which communication is easier between agents who are located closer together according to a social distance metric.

In contrast to these theories, in my theory the reputation effect makes relationships between members of different groups less profitable than relationships between members of the same group, even though there are no economically meaningful differences between members of different groups. The reputation effect appears endogenously in equilibrium, in

contrast to the exogenous differences that define primordial divisions. For this reason I refer to the divisions described by my model as "socially constructed" divisions.

Akerlof (1976) and Pęski and Szentes (2013) discuss a different kind of socially constructed division. These models are distinguished from my model by their information structure. In Akerlof and Pęski and Szentes, agents can observe something about with whom their partners have interacted in the past, with whom their partners' partners have interacted in the past, and so on to infinity. Akerlof and Pęski and Szentes use this information structure to construct an equilibrium in which an agent who interacts with a member of a different group is punished by her next partner, if her partner fails to punish then she is punished in turn by her next partner, and so on. In contrast in my model agents observe something about with whom their partners have interacted in the past, but that is all. In addition in my model there is no punishment for failing to punish a deviating agent. In section 3 I argue that the models of Akerlof and Pęski and Szentes represent a society in which there is a centralized institution that is specifically designed to gather the information and inflict the punishments necessary to support segregation. For this reason I refer to divisions described by Akerlof and Pęski and Szentes as "centralized" divisions. In contrast my model represents a society in which segregation is enforced without institutions specifically designed for the purpose, and so I refer to the divisions described by my model as "decentralized" divisions. I provide evidence from the anthropological literature that suggests that some Indian castes are decentralized while others are centralized.

The distinction between primordial and socially constructed groups yields insights about the origin and possible future of social division. Primordial divisions depend on some exogenous factor that creates a difference between members of different groups. For example, Michalopoulos (2012) describes a process in which a geographical barrier such as a mountain range divides a population, allowing the languages of the divided groups to diverge due to random drift. When the groups later recombine, the language barrier prevents them from interacting. In contrast, socially constructed divisions can appear even in the absence of any external cause, in a population of *ex ante* identical agents.

Different kinds of divisions are also likely to disappear in different ways. Primordial divisions depend on deep differences between members of different groups, and these differences change slowly if at all in response to changing economic conditions. Thus policy changes are unlikely to affect primordial divisions in the short term. In contrast, socially constructed divisions are an equilibrium outcome, and socio-economic parameters determine whether group division is a possible equilibrium. If socio-economic parameters change, socially constructed divisions may disappear suddenly, even in societies that have been segregated for thousands of years in the past. Thus even though the problems associated with social division, such as political conflict and violence, may appear intractable, we can have hope that with the right policies, it may be easier to ameliorate these problems than we think.

2 Model

2.1 Setup

Time is discrete, starts at period 0 and continues forever. A mass 1 of agents are born at the beginning of each period. Each agent is a member of one of G groups, and the mass of newborn agents from each group in each period is $1/G$. These groups are payoff irrelevant, but group membership is observable. Each agent has N relationship "slots". Newborn agents come into existence already matched with $N-1$ partners who are members of the same group. Thus each newborn agent has one relationship slot open. All agents have a fixed discount factor δ. In each period the following things happen:

1. Each agent with an open relationship slot pays a search cost c and is provisionally matched with another agent with an open slot. Agents are provisionally matched according to a uniform probability distribution over the set of agents with open relationship slots. More formally, as will be seen below an agent can be completely characterized by her group g and a what I call the agent's past match set, $\mathcal{H}$. There are a finite number of possible tuples $(g, \mathcal{H})$. The probability that an agent is provisionally matched with a partner with group and past match set $(g, \mathcal{H})$ is just the proportion of agents with group and past match set $(g, \mathcal{H})$ within the population of all agents with open relationship slots. An agent can search for at most one new partner in any period, even if she has more than one open relationship slot.

2. Provisionally matched agents observe their partners' groups and past match sets. Each agent may then choose to accept or reject the match. If either agent rejects the match, then the match is dissolved and both agents return to step 1. Otherwise a match forms and both agents continue to step 3.

3. All agents play a stage game with each of their partners, described below. The total payoff for each agent for the period is the sum of any the payoffs from the stage game in each relationship, minus any search costs.

4. For each matched pair of agents i and j, let a_i and a_j be the actions chosen in the stage game for that match. If $a_i \neq a_j$, then the match breaks up and both players begin the next period with an open relationship slot. Otherwise the match continues to the next period.

The fact that agents can have more than one relationship at a time is important for the model as it generates the possibility of externalities across matches. In particular, if an agent chooses to interact with a member of a different group, this can have consequences for her other concurrent relationships with members of her own group. This potential externality across relationships is important for the construction of my equilibrium.

It will turn out that on the equilibrium path matches never break up. This means that newborn agents are necessary to maintain a pool of agents with unfilled relationship slots, which in turn is necessary to define the expected cost of searching for a new partner and the expected cost of breaking up a relationship. An alternative would be to have a fixed set of agents, but to have relationships break up with some exogenous probability p. This alternative is more complicated, primarily due to the necessity of accounting for the possibility that two of an agents' relationships might break up simultaneously, but does not change the main results. The assumption that newborn agents are born matched with $N-1$ partners from their own groups also makes the presentation of the model simpler, again by ensuring that agents have at most one open relationship slot at a a time. This assumption can be justified by supposing that people are born with connections to family members and childhood friends and then must search for additional relationships as adults.

The stage game is as follows.[1] Both partners in the relationship simultaneously choose a stage game action $a \in [0, \infty)$. An agent's payoff is $\Pi(a, a')$, where the agent chooses action a and her partner chooses action a'. Define $v(a) = \Pi(a, a)$ and $d(a) = \Pi(0, a)$. I make the following assumptions on Π, v, and d:

Assumption 1. *1. For all $a > 0$ and all a', $\Pi(0, a') > \Pi(a, a')$.*

2. $v(a)$ is bounded.

3. $v(0) = d(0) = 0$

4. $v(a)$ and $d(a)$ are continuous, twice differentiable, and strictly increasing in a.

5. $v(a)$ is strictly concave in a and $d(a)$ is strictly convex in a.

Part 1 of the assumption states that 0 is the strictly dominant action in the stage game, which can be interpreted as a generalized prisoner's dilemma with a continuum of actions. If both players play a then both receive a payoff $v(a)$, and I will sometimes refer to this as the value of cooperation at level a. If one player plays a and the other plays 0, then the player who plays 0 gets $d(a)$, and I will sometimes refer to this as the value of cheating at level a. Part 2 is required to rule out Ponzi schemes, in which any level of cooperation can be attained through the promise of ever higher levels of cooperation in the future. Parts 3 through 5 imply that the temptation to cheat is small for a small, and that the temptation to cheat grows large as a gets large. These assumptions ensure that the solution to each agent's maximization problem is interior.

At the end of a period, a match breaks up if both partners to the match do not choose the same action in the stage game. The interpretation here is that agents must agree on a common level of cooperation a, and any deviation from this common level of cooperation is

[1] This stage game was first described in Ghosh and Ray (1996).

considered to be a violation of the agreement. By fixing the division of the surplus within each match, the assumption that agents must choose a common level of cooperation allows me to sidestep the issue of how to model bargaining within each relationship. Modelling bargaining in repeated games is very difficult and is beyond the scope of this paper. Miller and Watson (2013) provide one recent attempt at constructing such a model. The assumption that matches break up automatically after deviations can be justified as a behavioural response, if agents who are cheated become angry and refuse to work with the cheating partner in the future. A similar assumption appears in many other relational contracting models; see for example Levin (2003).

Each agent can observe her group and the group of any other agent with whom she is matched. Each agent can also observe the history of play within each current match, but she cannot observe the history of play in any match in which she does not participate. However, each agent can observe something about with whom each of her partners has matched in the past. Specifically, for each group g, an agent can observe whether any of her current partners have ever been matched with any agent in group g. Let $\mathcal{H}_i \subseteq \{1, ...G\}$ be the set of groups g such that agent i has been matched with a member of group g in the past. I refer to the set $\mathcal{H}_i$ as agent i's past match set. Note that for all groups g, if agent i is in group g then $g \in \mathcal{H}_i$, since agents are born matched to $N-1$ members of their own groups.

Let $a_{ij}^{\tau_{ij}}$ be the action played by player i in her match with player j in the τ_{ij}th period of the match. Then the history of player i's match with player j that has lasted for τ_{ij} periods is $h_{ij}^{\tau_{ij}} = \{(a_{ij}^1, a_{ji}^1), ..., (a_{ij}^{\tau_{ij}}, a_{ji}^{\tau_{ji}})\}$. Let h_i be the set of histories of all matches in which player i is currently engaged.

A (pure) strategy for agent i when matched or provisionally matched with partner j is a tuple $s_{ij}(h_{ij}^{\tau_{ij}}, g_i, g_j, \mathcal{H}_i, \mathcal{H}_j) = \{m_{ij}(g_i, g_j, \mathcal{H}_i, \mathcal{H}_j),\ a_{ij}(h_{ij}^{\tau_{ij}-1}, g_i, g_j, \mathcal{H}_j, \mathcal{H}_j)\}$. Here $m_{ij} \in \{A, R\}$ is agent i's decision to accept or reject a match with partner j observing agent j's group and past match set, and a_{ij} is the stage game action. A strategy s_i for player i is a set containing s_{ij} for all j such that $s_{ij} = s_{ij'}$ for all j, j'. That is, player i's strategy cannot depend on her partner's identity j, although player i's actions may depend on her partner's group and past match set g_j and $\mathcal{H}_j$. However, I allow agents to consider deviations that do depend on the identity of their partners. This allows for the possibility that an agent could deviate in one of her matches while remaining on the equilibrium path in her other contemporaneous matches. Note that I do not allow agent i to condition her action with partner j on the history of any prior match or on the history of any contemporaneous match with any partner $j' \neq j$. A strategy profile s is a set containing s_i for all i.

Suppose that an agent i with group and past match set $(g_i, \mathcal{H}_i)$ is matched with partners $j_1, ..., j_N$ with groups and past match sets $g_{j_1}, ..., g_{j_N}$ and $\mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}$. Let s be the strategy profile, and let s_{-i} denote the strategies played by all agents other than agent i, including agents $j_1, ..., j_N$. Suppose that the history of player i's current matches is h_i.

Then I denote player i's expected payoff by $EU_i[s_i, g_i, \mathcal{H}_i, s_{-i}, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}, h_i]$. When describing the renegotiation-proofness condition for equilibrium it will be helpful to explicitly denote the strategy of player j_1. In this case I denote player i's expected payoff by $EU_i[s_i, s_{j_1}, g_i, \mathcal{H}_i, s_{-i}, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}, h_i]$. Finally, denote the strategy for player i of playing strategy s_{ij} in matches with player j and strategy s_i in matches with all other players by $\frac{s_{ij}}{s_i}$.

2.2 Equilibrium Concept

An equilibrium of my model must satisfy two conditions. The first condition is an individual incentive compatibility condition. Let s^* be an equilibrium strategy profile. Let $\mathcal{E}^*$ be the set of combinations of group and past match history $(g, \mathcal{H})$ that an agent could have in a period if all agents are following strategy profile s^*. Then the first condition for an equilibrium is that s^* must satisfy

$$EU_i[s_i^*, g_i, \mathcal{H}_i, s_{-i}^*, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, \mathcal{H}_{j_N}, h_i] \geq EU_i[s_i, g_i, \mathcal{H}_i, s_{-i}^*, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, \mathcal{H}_{j_N}, h_i]$$

for all strategies s_i, for all $g_i, \mathcal{H}_i$, $g_{j_1}, \mathcal{H}_{j_1}$, and h_i, and for all $(g_{j_n}, \mathcal{H}_{j_n})$ such that $(g_{j_n}, \mathcal{H}_{j_n}) \in \mathcal{E}^*$ for all $n \geq 2$.

I comment briefly on the first equilibrium condition. Consider an agent i. The equilibrium condition states that an equilibrium strategy must be optimal for an agent i with any possible combination of group and past match set $(g_i, \mathcal{H}_i)$. An agent from group g_i could acquire past match set $\mathcal{H}_i$ through a sequence of deviations from the equilibrium strategy profile, and so the equilibrium condition requires that the equilibrium be robust to any number of past deviations by agent i. Not only is it possible that agent i could have deviated in the past, it is also possible that agent i could encounter a potential partner who as deviated in the past. Without loss of generality I suppose that the deviant partner is partner j_1. Thus, the equilibrium must be optimal for any possible combination of groups and past match sets $(g_{j_1}, \mathcal{H}_{j_1})$. However, the equilibrium condition does not require that the strategy profile be optimal for an agent who is matched with two or more partners, all of whom have deviated in the past. In addition, the equilibrium condition does not require that the strategy profile be optimal for an agent who expects future encounters with partners who have deviated in the past with positive probability. The idea here is that the equilibrium is robust to any sequence of deviations by any subset of agents with zero mass, but is not necessarily robust to the possibility that a positive mass of agents might deviate. If only a zero mass of agents have deviated in the past, then encounters between two or more agents, all of whom have deviated in the past, happen with probability zero, and each agent expects to encounter a partner who has deviated in the past with probability zero. It

seems plausible that a strategy profile that is robust to small numbers of deviations in this way could persist in a society, even if the equilibrium could be destroyed by a large number of simultaneous deviations. The equilibrium condition is similar to the norm equilibrium proposed by Okuno-Fujiwara and Postlewaite (1995), which is also robust to deviations by a mass zero set of agents but not to deviations by a positive mass of agents. Finally, I do not require that a strategy profile be optimal after an agent has deviated simultaneously in multiple relationships, which could lead to a situation in which an agent has fewer than N partners in a period. Since agents can search for only one partner in a given period, if it is not optimal for an agent to cheat in one relationship then it is also not optimal for the agent to cheat in multiple relationships simultaneously, regardless of the agent's continuation strategy. This fact, I argue, makes it unnecessary to account for the possibility that an agent might cheat in multiple relationships simultaneously.

The second condition is a renegotiation proofness condition. I adapt the condition from Ghosh and Ray (1996), and following their lead, I call the condition bilateral rationality. The condition the following:

There do not exist any g_i, $\mathcal{H}_i$, h_i, h_{j_1}, $g_{j_1}, ..., g_{j_N}$, $\mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}$, $g_{k_2}, ..., g_{k_N}$, $\mathcal{H}_{k_2}, ..., \mathcal{H}_{k_N}$, s_{ij_1}, $s_{j_1 i}$ such that $(g_{j_n}, \mathcal{H}_{j_n}) \in \mathcal{E}^*$ for all $n \geq 2$, $(g_{k_n}, \mathcal{H}_{k_n}) \in \mathcal{E}^*$ for all $n \geq 2$, and

$$EU_i \left[\frac{s_{ij_1}}{s_i^*}, \frac{s_{j_1 i}}{s_{j_1}^*}, g_i, \mathcal{H}_i, s_{-i}^*, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}, h_i \right] \geq$$
$$EU_i \left[s_i', \frac{s_{j_1}}{s_{j_1}^*}, g_i, \mathcal{H}_i, s_{-i}^*, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}, h_i \right]$$

for all s_i' and

$$EU_{j_1} \left[\frac{s_{j_1 i}}{s_{j_1}^*}, \frac{s_{ij_1}}{s_i^*}, g_{j_1}, \mathcal{H}_{j_1}, s_{-i}^*, g_i, g_{k_2}, ..., g_{k_N}, \mathcal{H}_i, \mathcal{H}_{k_2}, ...\mathcal{H}_{k_N}, h_{j_1} \right] \geq$$
$$EU_{j_1} \left[s_{j_1}', \frac{s_{ij_1}}{s_i^*}, g_{j_1}, \mathcal{H}_{j_1}, s_{-i}^*, g_i, g_{k_2}, ..., g_{k_N}, \mathcal{H}_i, \mathcal{H}_{k_2}, ...\mathcal{H}_{k_N}, h_{j_1} \right]$$

for all s_j'

and

$$EU_i\left[\frac{s_{ij_1}}{s_i}, \frac{s_{j_1 i}}{s_{j_1}}, g_i, \mathcal{H}_i, s^*_{-i}, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}, h_i\right] \geq$$
$$EU_i\left[s^*_i, s^*_{j_1}, g_i, \mathcal{H}_i, s^*_{-i}, g_{j_1}, ..., g_{j_N}, \mathcal{H}_{j_1}, ..., \mathcal{H}_{j_N}, h_i\right]$$

and

$$EU_{j_1}\left[\frac{s_{j_1 i}}{s^*_{j_1}}, \frac{s_{ij_1}}{s^*_{j_1}}, g_{j_1}, \mathcal{H}_{j_1}, s^*_{-i}, g_i, g_{k_2}, ..., g_{k_N}, \mathcal{H}_i, \mathcal{H}_{k_2}, ..., \mathcal{H}_{k_N}, h_{j_1}\right] \geq$$
$$EU_{j_1}\left[s^*_{j_1}, s^*_i, g_{j_1}, \mathcal{H}_{j_1}, s^*_{-i}, g_i, g_{k_2}..., g_{k_N}, \mathcal{H}_i, \mathcal{H}_{k_2}..., \mathcal{H}_{k_N}, h_{j_1}\right]$$

with at least one of the last two inequalities strict.

The bilateral rationality condition states that for any two matched agents i and j_1, it must not be possible for the agents to agree to deviate to a new strategy that satisfies the individual incentive compatibility condition for both of them and that provides both agents with higher utility (and at least one agent with strictly higher utility). The idea is that if two agents can communicate before they choose their actions, then they could agree to renegotiate to a new strategy if doing so would be mutually profitable, assuming that the new strategy is individually incentive compatible and hence credible for both agents. It is possible that either agent i or her partner may have deviated from the strategy profile in the past, and so the bilateral rationality condition must hold for agents i and j_1 with any possible combinations of group and past match set $(g_i, \mathcal{H}_i)$ and $(g_{j_1}, \mathcal{H}_{j_1})$. However, as before I do not require the bilateral rationality condition to hold for agents who are matched with two or more partners who have deviated in the past, or for agents who expect to be matched in the future with positive probability with partners who have deviated in the past, and I do not require the bilateral rationality condition to hold for agents who are matched with fewer than N partners.

As will become clear shortly, the bilateral rationality requirement plays a central role in motivating the introduction of the reputation effect that is the main contribution of this paper.

2.3 A Benchmark Equilibrium

I will begin my analysis by discussing a benchmark strategy profile in which agents do not condition their actions on their own or their partner's group membership or past match set. If the benchmark strategy profile is part of an equilibrium, I will refer to that equilibrium as a benchmark equilibrium.

In the benchmark strategy profile, every agent accepts every match regardless of group or past match history. In the stage game, each agent chooses action $\bar{a}_B$.

A benchmark strategy profile is an equilibrium if there are no profitable individual or

joint deviations. We must check that no agent can profit individually by cheating in any relationship, and also that no pair of matched agents can jointly profit by deviating to a higher level of cooperation that is individually incentive compatible for both agents. In principle, we also need to check that it is optimal for all agents to accept matches with all other members of the community. However, this last condition is trivial in the benchmark equilibrium, since all match partners are identical.

Because actions taken in one relationship do not affect any other relationship under the benchmark strategy profile, it is possible to analyse each relationship slot separately. Let V_B^u be the value that an agent expects to receive from an open relationship slot at the beginning of any period. Let V_B^m be the value that an agent expects to receive from a relationship slot that is filled at the beginning of a period. I also define V_B^f to be the expected value to each agent of having a filled relationship slot at the beginning of any future period. In the proof of proposition 1 it is helpful to distinguish V_B^f from V_B^m because agents may be able to affect V_B^m through renegotiation, but they cannot affect V_B^f. Bilateral rationality dictates that each pair of matched agents chooses the level of cooperation that maximizes their joint utility, subject to the constraint that no agent can profit individually by choosing to cheat. That is, V_B^m must satisfy:

$$V_B^m = \max_a v(a) \tag{1}$$

subject to the constraint

$$V_B^m \geq (1-\delta)d(a) + \delta V_B^u \tag{2}$$

Equation (1) says that an agent gets $v(a)+b$ from a match both in the current period and in all future periods. The constraint (2) is the individual incentive compatibility constraint. It states that the value of cooperating must be greater than the payoff that the agent receives from cheating. If the agent cheats she receives $d(a)+b$ in the current period and then gets the value of an empty relationship slot in the next period. The payoff to having an empty relationship slot V_B^u is defined by:

$$V_B^u = -(1-\delta)c + V_B^f \tag{3}$$

Equation (3) says that an agent with an empty relationship slot must pay the search cost in the current period before being matched with a new partner and receiving the payoff to that future match.

A benchmark equilibrium is a benchmark strategy profile such that V_B^m, V_B^u, and V_B^f satisfy equation (1) subject to (2) and equation (3), such that $\bar{a}_B$ maximizes (1) subject to (2), and such that $V_B^m = V_B^f$.

Define $\hat{a}$ to be the value of a that solves

$$\max_a v(a) - (1-\delta)d(a).$$

The following proposition provides conditions under which a benchmark equilibrium exists, and derives the level of cooperation in a benchmark equilibrium:

Proposition 1. *A benchmark equilibrium exists if and only if c satisfies*

$$c \geq \frac{1}{\delta}[d(\hat{a}) - v(\hat{a})]. \tag{4}$$

If a benchmark equilibrium exists, then the equilibrium level of cooperation $\bar{a}_B$ solves

$$d(\bar{a}_B) - v(\bar{a}_B) = \delta c \tag{5}$$

Omitted proofs are in appendix A.

The interpretation of the expression for the level of cooperation in the benchmark equilibrium is straightforward. If an agent cheats in the current period, her net gain in the period is the difference between the value of cheating $d(\bar{a}_B)$ and the value of cooperating $v(\bar{a}_B)$. The cost of cheating is that the cheating agent's match will break up, so that in the next period she will have to pay the search cost to find a new partner. Discounted for one period, this cost is δc. The maximum level of cooperation that can be sustained is the level of cooperation such that the net cost of cheating is equal to the net benefit. The bilateral rationality condition ensures that all agents will renegotiate up to the highest possible level of cooperation, so only the maximum sustainable level of cooperation is consistent with equilibrium.

I briefly discuss the intuition for the fact that no bilaterally rational equilibrium exists unless c is sufficiently large. I consider strategy profiles in which all agents choose the same level of cooperation every period. Since all agents accept all matches, any agent can cheat in her current relationship, break up the relationship at the end of the period, pay the search cost c, and find a new partner in the next period. Since all agents choose the same level of cooperation, the deviating agent will be able to cooperate at the same level in her new relationship as she did in the old relationship. Thus, if c is low, then the penalty for cheating in any given relationship is low, and so the common sustainable level of cooperation is low. However, if all agents are cooperating at some common low level, then any two matched agents can jointly deviate to a higher level of cooperation. This higher level of cooperation does not violate the individual incentive compatibility constraint, so long as only two agents are cooperating at the high level, because the penalty for breaking up this deviant relationship is high: if either agent breaks the relationship, both agents must go back to cooperating at the low common level of cooperation. Thus the individual incentive compatibility requirement rules out all strategy profiles except those strategy profiles with a low common level of cooperation, and the bilateral rationality requirement rules out strategy

profiles with a low common level of cooperation, so that there are no remaining equilibrium strategy profiles. As c gets larger, higher levels of cooperation become compatible with the individual incentive compatibility constraint, and for c sufficiently large there exist levels of cooperation that are high enough to satisfy the bilateral rationality requirement while still satisfying the individual incentive compatibility constraint.[2]

2.4 Motivating the Segregated Equilibrium

My goal is to construct an equilibrium that supports higher levels of cooperation than the benchmark equilibrium. I do this by constructing an equilibrium in which agents reject some matches, instead of accepting all matches as in the benchmark equilibrium. If agents reject some matches, then the expected cost of search for an unmatched agent is higher than in the benchmark equilibrium, and so the penalty for cheating and the level of cooperation that can be supported in each match are also higher.

The main barrier to constructing an equilibrium in which agents reject some potential matches is the bilateral rationality requirement. To build intuition for why bilateral rationality makes it difficult to construct such an equilibrium, consider the following strategy profile, which is a simplified version of the strategy profile considered by Eeckhout (2006), and which I will refer to as strategy profile E. Agents accept matches with members of their own group, and reject matches with members of any other group, regardless of past match histories. Within each match all agents choose action $\bar{a}_E$.

As in the benchmark equilibrium, under strategy profile E actions taken in one relationship slot do not affect the optimal action in any other relationship slot. Thus is is possible to analyse each relationship slot separately. Let V_E^m be the value to an agent from having a filled relationship slot in a period, and let V_E^u be the value to an agent from having an empty relationship slot in a period under strategy profile E. Under strategy profile E the composition of the pool of agents with unfilled relationship slots is strategically relevant. Fortunately the composition is easy to describe. If all agents follow strategy profile E, then in at the beginning of each period there are $\frac{1}{G}$ agents from each group in the pool of agents with unfilled relationship slots. Thus a searching agent meets a partner from her own group with probability $\frac{1}{G}$. Using this probability I can write expressions for V_E^m and V_E^u as follows:

$$\begin{aligned} V_E^m &= v(\bar{a}_E) \\ V_E^u &= -(1-\delta)c + \frac{1}{G}V_E^m + \frac{G-1}{G}V_E^u \end{aligned}$$

The first equation says that an agent who is matched cooperates forever at level $\bar{a}_E$. The

[2] A similar issue arises in Ghosh and Ray (1996), and the proof of proposition 1 draws on ideas from the proofs in that paper.

second equation says that an unmatched agent pays the search cost and is matched with a partner with probability $\frac{1}{G}$, and otherwise remains unmatched and must pay the search cost again.

Strategy profile E satisfies the individual incentive compatibility condition if

$$V_E^m \geq (1-\delta)d(\bar{a}_E) + \delta V_E^u$$

Rearranging these conditions yields that strategy profile E satisfies the individual incentive compatibility condition if

$$d(\bar{a}_E) - v(\bar{a}_E) \leq \delta G c$$

Comparing this expression to the expression defining the benchmark level of cooperation $\bar{a}_B$ yields the following:

Lemma 1. *If $G > 1$, then there exist values of $\bar{a}_E$ such that $\bar{a}_E > \bar{a}_B$ and such that strategy profile E satisfies the individual incentive compatibility condition.*

Higher levels of cooperation are individually incentive compatible under strategy profile E than in the benchmark equilibrium because agents expect to form matches with only $1/G$ of their potential partners under strategy profile E, while they expect to form matches with all of their potential partners in the benchmark equilibrium. Thus, the expected cost of breaking up a relationship is higher under strategy profile E than in the benchmark equilibrium, and so the individually incentive compatible level of cooperation is higher under strategy profile E than in the benchmark equilibrium.

We also have the following:

Lemma 2. *Strategy profile E is not an equilibrium because it does not satisfy the bilateral rationality condition.*

To see why strategy profile E is not bilaterally rational, suppose that $\bar{a}_E$ is such that strategy profile E satisfies the individual incentive compatibility condition, and consider two provisionally matched agents from different groups. The value to these provisionally matched agents of jointly deviating to accept the match is V_E^m, while the value of following the strategy profile and rejecting the match is V_E^u, where $V_E^m > V_E^u$. Moreover, this deviant relationship is individually incentive compatible for both agents. Thus there exists a mutually profitable and individually incentive compatible deviation from strategy profile E, and so strategy profile E does not satisfy the bilateral rationality requirement.

The problem with strategy profile E is that under the strategy profile relationships between members of different groups are just as profitable as relationships between members of the same group, and yet members of different groups do not interact. Intuitively it seems implausible to believe that people would consistently fail to seize opportunities for profitable

interaction in this way. The bilateral rationality requirement formalizes this intuition. A more plausible theory of group segregation would provide a reason why relationships between members of different groups are less profitable than relationships between members of the same group. In the next section I construct a strategy profile that contains just such a reason, and which therefore does satisfy the bilateral rationality requirement.

2.5 The Segregated Equilibrium

In this subsection I propose what I will call the segregated strategy profile. As before, if the segregated strategy profile is part of an equilibrium, I refer to the equilibrium as a segregated equilibrium. In the segregated equilibrium agents interact only with members of their own groups on the equilibrium path, which increases the cost of breaking up any match and thereby allows matched agents to support higher levels of cooperation than can be supported in the benchmark equilibrium. In addition, there is a reputational penalty for agents who interact with members of certain other groups. This reputational penalty makes interactions between members of different groups less profitable than interactions between members of the same group and thus ensures that the segregated strategy profile is bilaterally rational.

The segregated strategy profile is as follows. Groups are ranked in a hierarchy. I label the groups so that group 1 is ranked highest in the hierarchy and group G is ranked lowest. Thus $g > g'$ means that g is ranked below g'. An agent accepts matches with members of a group if and only if 1) the group is included in the agent's past match set, or 2) the group is of equal or higher rank to the agent's group. Formally, $m(g, \mathcal{H}, g', \mathcal{H}') = A$ if and only if $g' \leq g$ or $g' \in \mathcal{H}$. Matched agents choose a level of cooperation in each period that depends on the groups and past match histories of each of the partners to the relationship.

A segregated equilibrium is a segregated strategy profile for which there are no profitable individual or joint deviations. More specifically, a segregated equilibrium must satisfy four conditions. To cut down on notation I state these conditions informally. The conditions for an equilibrium are:

1. No agent can profit by cheating in any relationship.

2. No matched pair of agents can jointly profit by deviating to a higher level of cooperation that is also individually incentive compatible for both agents.

3. All agents prefer to accept matches with members of their own or higher ranking groups, or with members of groups that are in their past match sets, rather than rejecting those matches and continuing to search.

4. All agents prefer to reject matches with members of lower ranking groups that are not in their past match sets.

In order to describe a segregated equilibrium more formally, it will be useful to define the maximum level of cooperation that is individually incentive compatible for an agent under the segregated strategy profile. The level of cooperation that is individually incentive compatible for an agent depends on the probability that the agent will be able to find a new match in a period if her current match breaks up, which in turn depends on the probability of being provisionally matched with a member of each other group. Under the segregated strategy profile, the proportion of agents from each group in the pool of agents with empty relationship slots is $1/G$ in every period. Thus the maximum level of cooperation is individually incentive compatible for an agent depends only on the number of groups with whom the agent is expected to form matches, and not on the period or on which groups the agent is expected to match with. Let $\gamma(g, \mathcal{H})$ be the number of groups with whom an agent with group and past match set $(g, \mathcal{H})$ expects to form matches while following the segregated strategy profile. More formally, let $\gamma(g, \mathcal{H})$ be the number of groups g' such that $m(g, \mathcal{H}, g', \{g'\}) = A$ and $m(g', \{g'\}), g, \mathcal{H}) = A$.

Let $\bar{a}(\gamma)$ be the maximum level of cooperation that can be supported by an agent who expects to form matches with γ groups, assuming that the agent can achieve this level of cooperation in every match. Formally, $\bar{a}(\gamma)$ is defined by

$$v(\bar{a}(\gamma)) = (1-\delta)d(\bar{a}(\gamma)) + \delta V^u(\gamma) \tag{6}$$

Here $V^u(\gamma)$ is defined by

$$V^u(\gamma) = -(1-\delta)c + \frac{\gamma}{G}v(\bar{a}(\gamma)) + \frac{G-\gamma}{G}V^u(\gamma) \tag{7}$$

Equation (6) states that $\bar{a}(\gamma)$ is the level of cooperation at which an agent is just indifferent between cooperating forever or cheating and receiving the value $V^u(\gamma)$ of being unmatched in the next period. Equation (7) states that an unmatched agent pays the search cost c in the current period, and then is matched with another agent with probability $\frac{\gamma}{G}$. With probability $\frac{G-\gamma}{G}$ the agent does not find a match and must pay the search cost again.

Rearranging equations (6) and (7) yields that $\bar{a}(\gamma)$ is the solution to the following equation:

$$d(\bar{a}(\gamma)) - v(\bar{a}(\gamma)) = \delta\frac{G}{\gamma}c$$

Comparing this this equation to the equation defining the equilibrium level of cooperation in the benchmark equilibrium $\bar{a}_B$ shows that $\bar{a}(\gamma) > \bar{a}_B$ for all $\gamma < G$. An agent who expects to form matches with $\gamma < G$ groups expects to pay a higher search cost on breaking up a relationship and so can support a higher level of cooperation.

The following lemma is useful:

Lemma 3. *Suppose that* $c \geq \frac{1}{1-\delta}[d(\hat{a}) - v(\hat{a})]$. *Then the following inequalities hold:*

$$V^u(1) < V^u(2) < ... < V^u(G) < V^m(G) < V^m(G-1) < ... < V^m(1)$$

It is intuitive that $V^m(\gamma) = v(\bar{a}(\gamma))$ is strictly decreasing in γ. An agent with large γ expects to form matches with higher probability when unmatched and so an agent with large γ pays a lower cost in expectation for breaking up a match. Thus an agent with large γ can support a lower level of cooperation. It is also straightforward that $V^u(G) < V^m(G)$, since $V^u(G) = -(1-\delta)c + V^m(G)$. It is slightly less straightforward to show that $V^u(\gamma)$ is strictly increasing in γ. The intuition is that if $V^u(\gamma) \leq V^u(\gamma')$ for some $\gamma > \gamma'$, then it would be possible for two matched agents who expect to cooperate at level $\bar{a}(\gamma)$ in all of their future matches to renegotiate up to cooperating at level $\bar{a}(\gamma')$ in the present match. But the condition on the search cost implies that such renegotiation is impossible, using the reasoning from proposition 1.

I can now state a proposition characterizing the segregated equilibrium and the conditions under which a segregated equilibrium exists, as follows:

Proposition 2. *Fix* $v(\cdot)$, $d(\cdot)$, b, c, *and* δ, *and suppose that*

$$c \geq \frac{1}{1-\delta}[d(\hat{a}) - v(\hat{a})].$$

Then there exists $\underline{N}$ *such that for all* $N > \underline{N}$, *a segregated equilibrium exists. If a segregated equilibrium exists, then the equilibrium level of cooperation chosen by an agent with group and past match set* $(g, \mathcal{H})$ *matched with a partner with group and past match set* $(g', \mathcal{H}')$ *is* $\min\{\bar{a}(\gamma(g, \mathcal{H})), \bar{a}(\gamma(g', \mathcal{H}'))\}$.

To understand the intuition underlying proposition 2, suppose that $G = 2$. There are four possible combinations of group and past match set $(1, \{1\})$, $(1, \{1, 2\})$, $(2, \{2\})$, and $(2, \{1, 2\})$. We have $\gamma(1, \{1\}) = \gamma(2, \{2\}) = \gamma(2, \{1, 2\}) = 1$, and $\gamma(1, \{1, 2\}) = 2$. Agents from group 1 with past match set $\{1\}$ expect to interact only with members of their own group, because they expect to reject matches with members of group 2. Agents from group 2 also expect to interact only with members of their own group, not because they expect to reject members of group 1 but because they expect to be rejected by members of group 1. Finally, agents from group 1 with past match set $\{1, 2\}$ expect to interact with members of both groups.

A necessary condition for the bilateral rationality condition to be satisfied is that all agents cooperate at the highest level that is individually incentive compatible for both partners to the match. That is, the level of cooperation in any match involving a partner with group and past match set $(1, \{1, 2\})$ is $\bar{a}(2)$, and the level of cooperation in any other match is $\bar{a}(1)$. Likewise the value of being matched for an agent in a match in involving a partner with group and past match set $(1, \{1, 2\})$ is $V^m(2)$, while the value of being matched

for any other agent is $V^m(1)$. In order for the bilateral rationality condition to be satisfied the search cost c must also be sufficiently large, for the same reason that the search cost must be sufficiently large for a benchmark equilibrium to exist.

Now consider a provisional match between an agent with group and past match set $(1, \{1\})$ and another agent with group and past match set $(2, \{2\})$. If the match forms, the past match set of both agents will change to $\{1, 2\}$. Both partners must decide whether or not to accept the match. Consider first the choice faced by the group 2 agent. If the agent rejects the match, she will get value $V^u(1)$ from the relationship slot, while if she accepts the match she will get value $V^m(2)$ from the relationship slot. In either case, the agent gets value $V^m(1)$ from her other $N-1$ relationship slots. Since $V^m(2) > V^u(1)$ by lemma 3, the group 2 agent prefers to accept the match.

On the other hand, consider the decision of the group 1 agent. Like the group 2 agent, the group 1 agent gets value $V^m(2)$ from the relationship slot if she accepts the match and value $V^u(1)$ if she rejects the match. However, if she accepts the match the value of all of her other relationships will fall to $V^m(2)$ from $V^m(1)$. Thus the group 1 agent's total payoff from accepting the match is $NV^m(2)$ and her total payoff from rejecting the match is $V^u(1) + (N-1)V^m(1)$. Since $V^m(1) > V^m(2)$, for N sufficiently large the group 1 agent prefers to reject the match.

Thus for N sufficiently large all of the conditions for the existence of a segregated equilibrium are met. The central idea is that it if an agent has interacted only with members of her own group in the past, then it is believed that the agent will continue to interact only with members of her own group in the future. If an agent chooses to interact with a member of a lower-ranking group, then it is believed that the agent will continue to form matches with members of the lower-ranking group in the future. Thus an agent who has accepted a match with members of lower ranking groups can support lower levels of cooperation in each relationship. Importantly, this reputational penalty affects all of the agent's relationships, not just her relationship with the member of the lower-ranking group. If the agent has a sufficiently large number of relationships then she prefers to reject matches with members of lower ranking groups in order to avoid this reputational penalty to her other relationships. In contrast, accepting a match with a member of a higher ranking group does not generate a reputational penalty, and so agents are willing to accept matches with members of higher-ranking groups.

I conclude this section by comparing welfare under the segregated equilibrium and under the benchmark equilibrium. The lifetime utility of a newborn agent under the segregated equilibrium is

$$W_S = V^u(1) + (N-1)V^m(1)$$

Lifetime utility for a newborn agent under the benchmark equilibrium is

$$W_B = V^u(G) + (N-1)V^m(G)$$

Since $V^m(1) > V^m(G)$, we have

Corollary 1. *There exists $\underline{N}$ such that for all $N > \underline{N}$, $W_S > W_B$.*

The segregated equilibrium features both higher levels of cooperation and higher search costs than the benchmark equilibrium, but for suffiently large N the increased value of cooperation outweighs the extra search costs, increasing total welfare. This result provides a reason to believe that the segregated equilibrium would be selected over the benchmark equilibrium.

2.6 Equilibrium Selection and the Hierarchy

I have shown that under the right conditions there exists an equilibrium that supports higher levels of cooperation that the benchmark equilibrium by preventing agents from interacting with members of different groups on the equilibrium path. The question remains whether there exist other equilibria in which agents interact only with members of their own groups, perhaps enforced by some mechanism other than the reputation mechanism at the heart of the segregated equilibrium. In particular the group hierarchy may seem like an ad hoc addition to the segregated strategy profile, raising the question of whether it is possible to construct an equilibrium without the hierarchy. In this section I argue that the segregated equilibrium, including the hierarchy, is in fact the most natural way to support high levels of cooperation in the environment that I describe.

Consider first the case where $G = 2$. In this case we have the following:

Proposition 3. *Suppose that $G = 2$, suppose that a segregated equilibrium exists, and suppose that the conditions for the existence of a segregated equilibrium are satisfied with strict inequality. Then the segregated equilibrium is the unique equilibrium in which agents interact only with members of their own groups on the equilibrium path.*

The intuition is as follows. The minimum value that an agent can get from a match in any equilibrium is $V^m(G)$. If a segregated equilibrium exists then $V^m(G) > V^u(\gamma)$ for all γ, so an agent always prefers to accept a match if doing so does not affect the level of cooperation that she can achieve in her other relationships. Therefore if members of groups 1 and group 2 do not interact on the equilibrium path, then a match between a member of group 1 and group 2 must affect either the group 1 agent's other relationships, or the group 2 agent's other relationships, or both. Without loss of generality, suppose that a match between a member of group 1 and group 2 affects the group 1 agent's other relationships. A match between a member of group 1 and group 2 could affect the group 1 agent's other relationships by making the agent's other partners believe that the group

1 agent will continue to form matches with group 2 agents in the future. But this belief is rational only if members of group 2 are willing to accept matches with members of group 1. Thus it must be the case that members of group 1 reject matches with members of group 2 while members of group 2 accept matches with members of group 1 on the equilibrium path, and that members of group 1 who have accepted matches with members of group 2 in the past continue to accept such matches in the future. This is just the segregated equilibrium.

With three or more groups, the segregated equilibrium is no longer unique. For example, with three groups there may be an equilibrium that is a cycle: members of group 1 reject matches with members of group 2, members of group 2 reject matches with members of group 3, and members of group 3 reject matches with members of group 1. Other patterns are possible with larger numbers of groups. I omit a complete classification of these equilibria as the classification quickly becomes complex.

Despite the non-uniqueness of the segregated equilibrium for $G > 2$, proposition 3 provides some reason to believe that the hierarchy that is part of the segregated equilibrium is a natural feature of segregated societies. This result corresponds nicely to the evidence presented in the introduction that the relationships between Indian castes are in fact hierarchical. The hierarchy is not due to intrinsic differences between groups but instead appears endogenously in equilibrium between groups that are *ex ante* identical. Thus just like the division of society into groups, the hierarchical relationships between groups are "socially constructed".

3 Centralized and Decentralized Segregation

So far I have developed a theory of social division in which members of different groups do not interact with each other due to a reputation effect. In this section I step back from formal modelling and return to considering whether my theory corresponds well to qualitative descriptions of real groups from the anthropological literature. Once again I use the Indian caste system as a case study. In the introduction I described some broad features of the Indian caste system that seem to match some of the features of my model. In this section I point out some more subtle features of the model. I contrast these features with the models in two other papers, Akerlof (1976) and Pęski and Szentes (2013) (henceforth APS). APS also feature agents who are members of payoff-irrelevant groups who search over a community to find relationship partners. The relationships are not modeled as cooperative, and so the models are not completely comparable to mine. However, APS do feature a reputation effect that is similar to the effect in my model. In APS, as in my model, agents are punished for interacting with members of different groups, and this punishment reduces the amount of interaction between members of different groups in equilibrium. However, the details of the mechanism are different, and I argue that the differences between my model

and APS can be seen in anthropological accounts of institutions in different Indian castes.

The first difference between my model and APS is in the information structure of the community. In my model, in the segregated equilibrium agents can observe some information about their partners' previous actions. Specifically, agents have some information about with whom their partners have interacted in the past. The information structure in APS is related but is much richer. In APS agents have information about with whom their partners have interacted with in the past, with whom their partners' partners have interacted with in the past, and so on to infinity.[3] In my opinion it is implausible that people could get information about with whom their partners' partners'... partners have interacted in the past through gossip or other informal processes of information sharing. However, it may be more plausible to believe that agents could have the detailed information described by APS if there is a centralized institution specifically devoted to collecting and disseminating this information. In fact the Indian caste system does feature just such an institution, the caste panchayat. The panchayat is an assembly of caste members in a village that provides governance for the caste and makes decisions about the caste rules, including rules about interaction with other castes. In castes where the panchayat collects detailed information each person's interaction partners and disseminates this information to the community, it may be plausible to model people as observing not just with whom their partners have interacted in the past but also with whom their partners' partners have interacted and so on. In contrast, in castes where the panchayat either does not exist or plays a minimal role in information collection, it seems more plausible to model people as observing with whom their partners have interacted in the past but nothing more. For this reason I describe my model as a model of "decentralized" socially constructed division, in which there is no specialized institution that collects information about community interactions, while I say that APS describes a "centralized" socially constructed division.

A second difference between my model and APS is in the consequences for failing to inflict the punishments dictated by the equilibrium strategy profile. In my model, agents can be punished for deviating from the equilibrium strategy profile by interacting with members of lower ranking groups. The punishment is that deviating agents are supposed to cooperate at a lower level in their future relationships. However, suppose that for some reason some agent fails to inflict the punishment and instead cooperates at a high level with a partner who is supposed to be in the punishment phase. In my model there is no further punishment for this deviation. Indeed it would not be possible for agents to inflict such a further punishment because they do not know whether their partners have in fact inflicted the punishments that the strategy profile tells them to inflict. In contrast, in APS agents do know whether their partners have inflicted the punishments that they are supposed to inflict, and agents are punished for failing to punish, for failing to punish failure to punish,

[3]This kind of infinite regress appears frequently in the community enforcement literature. See, for example, Kandori (1992) and Okuno-Fujiwara and Postlewaite (1995).

and so on. Again, it seems implausible that this complex system of punishments could be sustained in a decentralized way. However, if there is a centralized institution specifically designed to keep track of who needs to be punished and for what, then it may be possible to maintain such a system.

A third difference between my model and APS is in the severity of punishments. In my model there is no punishment for failure to punish and so it must be incentive compatible and bilaterally rational for agents to inflict punishments even though their decision to punish or not has no effect on their payoffs in any other relationships. This means that punishments in my model must be relatively mild. In particular the most severe possible punishment, withdrawing all cooperation from a deviating agent, is not bilaterally rational and so is not used. Instead agents who deviate are able to support levels of cooperation that are lower but still positive. In contrast, in APS there are punishments for failure to punish. This means that it is possible to support more severe punishments in equilibrium.

The anthropological literature offers a wealth of case studies that allow us to compare the predictions of my model and the APS models. I consider two differing accounts of caste rules in India. An example where my model performs well comes from Hayden (1983). Hayden describes rules in the Nandiwalla caste in the state of Maharashtra. Among the Nandiwallas a person who has broken the caste rules is said to be *eli*. A person who has become *eli* can appear before the caste panchayat and be reinstated in the community by paying a fine. However, the panchayat does not seem to play a role in distributing information about who is *eli* and who is not. Hayden describes the process of becoming *eli* as follows: "*Eli* is not a status that is *imposed* on a person for his actions. Rather it is an automatic reaction to the fact that one automatically becomes polluted by an improper act.... It does not have to be pronounced by anyone." This seems to correspond to the idea that information among the Nandiwallas spreads in a decentralized and non-purposeful way.

Regarding the consequences of being *eli*, Hayden writes:

"The [Nandiwallas] say that they 'won't give even fire' to one who is *eli*. However, there is a certain literal quality to this pronouncement. They won't give him fire, but they will give him matches. They won't take food with him, but they will certainly drink liquor and take *pa:n* with him. One should not quarrel with someone who is *eli*, but the latter may argue in panchayat. What seems to happen is that, although certain specific commensal activities with other caste members are limited for one who is *eli*, most aspects of his life remain unchanged. He still puts his tent in the same place in both the large triennial encampment and in smaller camps on the road. People come to visit, and he can reciprocate. In most ways, life goes on normally.

In addition to those mentioned above, the activities in which an 'outcast's' [*sic*] participation is restricted involve business, religious ceremonies, and the marriage of his children. In the first category, one should not enter into any business arrangement with one who is *eli*. In the second, those who are not fully in caste cannot participate in most group religious ceremonies. It is the third category that is potentially the most serious. If one is in caste suspension, he cannot arrange marriages for his young children, and other families should

not honor previous arrangements by accepting his daughter or sending their own so long as he is *eli*. However, if someone does honor a marriage agreement, the amount he is charged is usually small."

This description seems to correspond to the ideas in my model that there is no punishment for failure to punish and that punishments for deviators are relatively mild. Nandiwallas who have become *eli* are limited in some of their interactions with other caste members, but they are not cut off from all interaction. If someone does interact in a supposedly prohibited way, for example by honoring a marriage agreement with an *eli* family, the punishment for this failure to punish is only a nominal fine.

A quite different account of caste rules comes from Majumdar (1958), who describes a dispute in the Chamar caste in the state of Uttar Pradesh. His description of the case is as follows:

"Even if a person gives food or water to an outcaste, or invites him for a smoke, without knowing the stigma attached to the recipient of his kindness, the unwitting offender also relinquishes his membership of the caste.... An instance of this occurred in May 1954, when K-Chamar of Bijapur village visited B-Chamar of Mohana. K-Chamar had been, for some reason or other expelled from his caste by the Chamar *biradari* [the local word for the panchayat] of Bijapur. He came to Mohana without letting anyone know of the disgrace, and B-Chamar as is the custom treated his guest very hospitably, and they took their midday meals together. Soon it was known that K-Chamar was an outcaste. Consequently B-Chamar was declared an outcaste by the Chamar caste-panchayat of Mohana."

This account contrasts with the account of Nandiwallas on all three of the dimensions described above. First, the caste panchayat plays a major role in disseminating the information needed to inflict the required punishments. Without having heard the proclamations of the panchayat, B-Chamar did not know that he was supposed to punish K-Chamar and so he did not do so. Second, people can be punished not just for breaking the caste rules but also for failure to punish others who have broken the caste rules. B-Chamar is punished for failure to punish K-Chamar, and B-Chamar's punishment is just as severe as K-Chamar's. Third, the punishment for each crime is withdrawal of nearly all interaction from the guilty parties. It seems, then, that the Chamars have a centralized system of segregation perhaps better described by APS than by my model.

In short, then the Indian caste system seems to contain examples both of decentralized segregation, described by my model, and centralized segregation, better described by APS. It would be interesting to study the development of these institutions over time. My intuition is that decentralized social divisions appear first, and then later become centralized as the level of cooperation and hence the degree of segregation required by the community increases. I know of no study on this topic, however.

4 The Origin and Future of Social Division

In this paper I have developed a model of social divisions in which a reputation effect prevents interactions between members of different groups in equilibrium, even though those interactions would be just as profitable as interactions between members of the same group in the absence of the reputation effect. My theory is a theory of "socially constructed" divisions, as the distinction between different groups is an endogenous feature of the equilibrium and not due to any fundamental difference between members of different groups. A competing theory is that members of different groups do not interact because some difference between members of different groups makes interactions between members of different groups less profitable than interactions between members of the same group. Theories of this type are theories of "primordial" division. Primordial divisions are created by some exogenous shock that generates a difference between different members of society. In contrast, socially constructed divisions can appear even in the absence of any external cause. Because socially constructed divisions do not depend on any external cause, the location of socially constructed group boundaries can be arbitrary.

Changes in socio-economic parameters affect primordial and socially constructed divisions in different ways. Primordial divisions are based on deep differences between people that respond only slowly to changing socio-economic conditions. In contrast socially constructed divisions are an equilibrium phenomenon that can disappear quickly if socio-economic parameters change in a way that destroys the segregated equilibrium. One kind of socio-economic change that could affect segregation is an improvement in formal contracting institutions. In order to model this effect, suppose that there is an additional step $2'$ between steps 2 and 3 in the sequence of events that happens each period from section 2.1. In step $2'$, agents can choose whether to enter into a cooperative or a non-cooperative relationship. If either agent chooses to enter into a non-cooperative relationship, then both agents receive a fixed payoff b instead of playing the stage game. If both agents choose to enter the cooperative relationship, then agents continue to step 3 to play the stage game as before. In this environment, the segregated strategy profile mandates that agents choose the cooperative relationship in all relationships rather than taking the fixed payoff b. However, the addition of the fixed payoff creates a new way for agents to deviate, by choosign the fixed payoff instead of the cooperative relationship. In order for an equilibrium to exist, it must be optimal for agents to choose the cooperative relationship in each period.

The fixed payoff b represents the payoff to a short-term relationship whose value does not depend on the intertemporal incentives required for cooperation. Many kinds of relationships can be short-term or long-term. For example, consider a credit relationship. The parameter b might represent the value of a collateralized loan. The collateral allows the lender to enforce repayment even in the absence of repeat lending. In contrast, $v(a)$ might represent the value of an uncollateralized loan, which the borrower repays in order to retain access to

future loans. Similarly, in a trade relationship, b might represent the value of trade in goods of observable quality, for which there is no possibility of cheating. In contrast, $v(a)$ might represent the value of trade in goods of unobservable quality, where the traders refrain from cheating each other because of the value of future trade.

In this environment, an increase in b represents an improvement in society's formal contracting institutions, which makes it possible to get value from a wider range of relationships even in the absence of long-term incentives. To see the effect of an increase in b, consider again the case where $G = 2$, and consider two provisionally matched agents, one with group and past match set $(1, \{1\})$ and the other with group and past match set $(2, \{2\})$. In order for a segregated equilibrium to exist, it must be the case that the group 1 agent prefers to reject the match. Following the previous analysis, if the group 1 agent follows the segregated strategy profile and rejects the match, she gets value at most $V^u(1) + (N-1)V^m(1)$. If she accepts the match, then she chooses either the cooperative or the non-cooperative relationship, depending on which is more profitable, and gets value $N \max\{b, V^m(2)\}$. For b sufficiently large, the value of accepting the match is greater than the value of following the segregated strategy profile. Thus we have the following corollary to proposition 2:

Corollary 2. *Fix $v(a)$, $d(a)$, c, δ and N. There exists $\underline{b}$ such that for all $b \geq \underline{b}$, no segregated equilibrium exists.*

There is some empirical evidence that improvements in formal contracting do in fact lead to the breakdown of segregation. Munshi and Rosenzweig (2006) study the effects of increasing economic integration with the outside world on castes in Mumbai. Over the period they study, increasing trade opportunities increased the relative value of formal sector employment as compared to informal sector employment, which in the context of my model could be thought of as an increase in b. Munshi and Rosenzweig show that the percentage of people marrying outside of their castes increased dramatically as formal sector employment opportunities improved. In the context of my model, this could be interpreted as a breakdown of the segregated equilibrium.

The breakdown of socially constructed divisions allows people to take advantage of the full range of possible relationships available to them, and it may also help to ameliorate other problems caused by division such as political conflict and violence. At the same time, the breakdown of these divisions is likely to lead to the loss of traditional community values and the high levels of cooperation that they entail. Finding a balance between these conflicting sets of values is likely to be one of the key issues for many countries as they move forward in the process of development.

References

[1] Akerlof, George (1976), "The Economics of Caste and of the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics* 90:4, 599-617.

[2] Akerlof, George and Rachel E. Kranton (2000), "Economics and Identity," *Quarterly Journal of Economics* 115:3, 715-753.

[3] Alesina, Alberto, Reza Baqir, and William Easterly (1999), "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics* 114:4, 1243-1284.

[4] Anderson, Siwan (2011), "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics* 3:1, 239-263.

[5] Becker, Gary S. (1957), *The Economics of Discrimination*, Chicago: University of Chicago Press.

[6] Bernheim, B. Douglas, and Debraj Ray (1989), "Collective dynamic consistency in repeated games," *Games and Economic Behavior* 1:4, 295-326.

[7] Bisin, Alberto and Thierry Verdier (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics* 115:3, 955-988.

[8] Chandra, Kanchan (ed.) (2012), *Constructivist Theories of Ethnic Politics*, Oxford: Oxford University Press.

[9] Cornell, Brad and Ivo Welch (1996), "Culture, Information and Screening Discrimination," *Journal of Political Economy* 104:3, 542-571.

[10] Dixit, Avinash K. (2003), "Trade Expansion and Contract Enforcement," *Journal of Political Economy* 111:6, 1293-1317.

[11] Eeckhout, Jan (2006), "Minorities and Endogenous Segregation," *Review of Economic Studies* 73:1, 31-53.

[12] Farrell, Joseph and Eric Maskin (1989), "Renegotiation in repeated games," *Games and Economic Behavior* 1:4, 327-360.

[13] Fearon, James D. and David D. Laitin (1996), "Explaining Interethnic Cooperation," *American Political Science Review* 90:4, 715-735.

[14] Genicot, Garance and Debraj Ray (2003), "Endogenous Group Formation in Risk-Sharing Arrangements," *Review of Economic Studes* 70:1, 87-113.

[15] Ghosh, Parikshit and Debraj Ray (1996), "Cooperation in Community Interaction Without Information Flows," *Review of Economic Studies* 63:3, 491-519.

[16] Greif, Avner (1993), "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review* 83:3, 525-548.

[17] Grimard, Franque (1997), "Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire," *Journal of Development Economics* 53:2, 391-422

[18] Habyarimana, James, Macartan Humphreys, Dan Posner, and Jeremy Weinstein, "Why Does Ethnic Diversity Undermine Public Goods Provision? An Experimental Approach," *American Political Science Review* 101:4, 709-725.

[19] Hayden, Robert M (1983),"Excommunication as Everyday Event and Ultimate Sanction: The Nature of Suspension from an Indian Caste", *Journal of Asian Studies* 42:2, 291-307.

[20] Kandori, Michihiro (1992), "Social Norms and Community Enforcement," *Review of Economic Studies* 59:1, 63-80.

[21] Lazear, Edward P. (1999), "Culture and Language," *Journal of Political Economy* 107:S6, S95-S126.

[22] Levin, Jonathan (2003), "Relational Incentive Contracts," *American Economic Review* 93:3, 835-857.

[23] Majumdar, D.N. (1958), *Caste and Communication in an Indian Village*, Asia Publishing House: Bombay.

[24] Mayer, Adrian C. (1960), *Caste and Kinship in Central India*, Berkeley: University of California Press.

[25] Mazzocco, Maurizio and Shiv Saini (2012), "Testing Efficient Risk Sharing with Heterogeneous Risk Preferences," *American Economic Review* 102:1, 428-468.

[26] Michalopoulos, Stelios (2012), "The Origins of Ethnolinguistic Diversity," *American Economic Review* 102:4, 1508-1539.

[27] Miguel, Edward and Mary Kay Gugerty (2005), "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics* 89:11-12, 2325-2368.

[28] Miller, David and Joel Watson (2013), "A Theory of Disagreement in Repeated Games with Bargaining", *Econometrica* 81:6, 2303-2350.

[29] Munshi, Kaivan and Mark Rosenzweig (2006), "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review* 96:4, 1225-1252.

[30] Munshi, Kaivan and Mark Rosenzweig (2009), "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth," *mimeo*, Brown University.

[31] Okuno-Fujiwara, Masahiro and Andrew Postlewaite (1995), "Social Norms and Random Matching Games," *Games and Economic Behavior* 9:1, 79-109.

[32] Pęski, Marcin and Balazs Szentes (2013), "Spontaneous Discrimination," *American Economic Review*.

[33] Tabellini, Guido (2008), "The Scope of Cooperation: Values and Incentives," *Quarterly Journal of Economics* 123:3, 905-950.

[34] Topkis, Donald M. (1998), *Supermodularity and Complementarity*, Princeton: Princeton University Press.

A Proofs

A.1 Proof of Proposition 1

Plugging equation (3) into the constraint (2) and equation (1) and rearranging yields

$$V_B^m = \max_a v(a) \tag{8}$$

subject to

$$V_B^f \leq \frac{1}{\delta}[v(a) - (1-\delta)d(a)] + (1-\delta)c \tag{9}$$

Recall that $\hat{a}$ was defined as the value of a that solves

$$\max_a v(a) - (1-\delta)d(a).$$

Since v is strictly concave and d is strictly convex, there exists a finite value of a that maximizes $\hat{a}$. Since $\hat{a}$ has a maximum value, there exists $\hat{V}_B^f$ such that the constraint (9) can be satisfied for $a \geq 0$ if and only if $V_B^f \leq \hat{V}_B^f$, with $\hat{V}_B^f$ defined by

$$\hat{V}_B^f = \frac{1}{\delta}[v(\hat{a}) - (1-\delta)d(\hat{a})] + (1-\delta)c. \tag{10}$$

Now, define a function $\phi(x)$ by

$$\phi(x) = \max_a v(a) \tag{11}$$

subject to

$$x \leq \frac{1}{\delta}[v(a) - (1-\delta)d(a)] + (1-\delta)c \tag{12}$$

Any fixed point of ϕ is a benchmark equilibrium. However, notice that ϕ is not well-defined for all x, since for $x > \hat{V}_B^f$ there is no $a \geq 0$ that satisfies (12). Since v and d are continuous and differentiable, ϕ is continuous and differentiable. By the envelope theorem,

$$\frac{\partial \phi}{\partial x} = \frac{\delta p}{1-\delta} - \psi < 1 \tag{13}$$

where $\psi > 0$ is the Lagrange multiplier on the constraint (12). Finally, $\phi(-\delta pc)$ is well-defined and $\phi(-\delta pc) \geq -\delta pc$. Since ϕ is continuous, $\frac{\partial \phi}{\partial x} < 1$, $\phi(-\delta pc)$ is well defined and $\phi(-\delta pc) \geq -\delta pc$, $\phi(V^f)$ has exactly one fixed point if and only if

$$\phi(\hat{V}^f) \leq \hat{V}^f \tag{14}$$

Plugging in the expression for $\hat{V}^f$ from (10) into (14) and rearranging yields the condition that a benchmark equilibrium exists if and only if

$$c \geq \frac{1}{\delta}[d(\hat{a} - v(\hat{a})].$$

This completes the proof.

A.2 Proof of Lemma 3

Since $\bar{a}(\gamma)$ is decreasing in γ and since $V^m(\gamma) = v(\bar{a}(\gamma))$, it is immediate that if $\gamma < \gamma'$ then $V^m(\gamma) > V^m(\gamma')$. Since $V^u(G) = -(1-\delta)c + V^m(G)$, $V^u(G) < V^m(G)$. Thus it only remains to be shown that if $\gamma < \gamma'$ then $V^u(\gamma) < V^u(\gamma')$.

Define $\phi(x, \gamma)$ by

$$\phi(x, \gamma) = \max_a v(a)$$

subject to the constraint

$$x \leq \frac{1}{\delta}[v(a) - (1-\delta)d(a)] + (1-\delta)\frac{G}{\gamma}c$$

From the proof of proposition 1, $\phi(x, \gamma)$ has a fixed point for all γ such that $1 \leq \gamma \leq G$ if and only if

$$c \geq \frac{1}{1-\delta}[d(\hat{a}) - v(\hat{a})] \tag{15}$$

Inspection of the definition of $V^m(\gamma)$ shows that if $\phi(x)$ has a fixed point, then $V^m(\gamma)$ is the fixed point of $\phi(x)$. Thus if c satisfies the condition above, then, rearranging the constraint in the definition of $\phi(x, \gamma)$, $\bar{a}(\gamma)$ solves

$$\max_a v(a)$$

subject to

$$v(a) \geq (1-\delta)d(a) + \delta V^u(\gamma)$$

The solution to the previous problem is decreasing in $V^u(\gamma)$, and $\bar{a}(\gamma)$ is decreasing in γ, which implies that $V^u(\gamma)$ must be increasing in γ, completing the proof.

A.3 Proof of Proposition 2

The definition of $\bar{a}(\gamma)$ ensures that the segregated equilibrium levels of cooperation satisfy the individual incentive compatibility condition. From the proof of lemma 3, we have that $\bar{a}(\gamma)$ solves

$$\max_a v(a)$$

subject to

$$v(a) \geq (1-\delta)d(a) + \delta V^u(\gamma)$$

and so the bilateral rationality condition is satisfied. Thus it only remains to show

Consider an agent with group and past match set $(g, \mathcal{H})$ who is provisionally matched with a partner with group and past match set $(g', \mathcal{H}')$. Suppose that either $g' \leq g$ or $g' \in \mathcal{H}$. Then $\gamma(g, \mathcal{H}) = \gamma(g, \mathcal{H} \cup \{g'\})$, and so accepting the match does not affect the level of cooperation that can be achieved in the agent's other relationships. Thus the total value to the agent of accepting the match is at least $V^m(G)+(N-1)V^m(\gamma(g, \mathcal{H}))$. The total value of rejecting the match is at most $V^u(\gamma(g, \mathcal{H}))+(N-1)V^m(\gamma(g, \mathcal{H}))$. Since $V^m(G) > V^u(\gamma)$ for all γ by lemma 3, the agent prefers to accept the match.

Now suppose that $g' > g$ and that $g' \notin \mathcal{H}$. Then $\gamma(g, \mathcal{H} \cup \{g'\}) = \gamma(g, \mathcal{H}) + 1$. So the total value to the agent of accepting the match is at most $NV^m(\gamma(g, \mathcal{H}) + 1)$, while the total value to the agent of rejecting the match is at least $V^u(\gamma(g, \mathcal{H}))+(N-1)V^m(\gamma(g, \mathcal{H}))$. Since $V^m(\gamma(g, \mathcal{H})) > V^m(\gamma(g, \mathcal{H}) + 1)$, for N sufficiently large the agent prefers to reject the match.

A.4 Proof of Proposition 3

I prove the proposition by contradiction. First suppose that $m(1, \{1\}, 2, \{2\}) = m(2, \{2\}, 1, \{1\}) = A$. But then agents would accept matches with members of different groups on the equilibrium path, contradicting the assumption.

Now suppose that $m(1, \{1\}, 2, \{2\}) = m(2, \{2\}, 1, \{1\}) = R$. By assumption this means that agents with group and past match set $(1, \{1\})$ strictly prefer to accept matches with partners with group and past match set $(2, \{2\})$. Since the value to a group 1 agent of accepting a match with an partner with group and past match set $(2, \{2\})$ is the same as the value of accepting a match with an partner with group and past match set $(2, \{1, 2\})$, we must have $m(1, \{1\}, 2, \{1, 2\}) = R$. So $\gamma(1, \{1, 2\}) = 1$. A similar argument shows that $\gamma(2, \{1, 2\}) = 1$. So an agent from group 1 who accepts a match with a member of group 2 gets value $NV^m(1)$, while by rejecting the match the agent gets $V^u(1) + (N-1)V^m(1)$. Since $V^u(1) > V^m(1)$, the agent prefers to accept the match. But this contradicts the

assumption that $m(1, \{1\}, 2, \{2\}) = R$.

Thus either $m(1, \{1\}, 2, \{2\}) = R$ and $m(2, \{2\}, 1, \{1\}) = A$ or $m(1, \{1\}, 2, \{2\}) = A$ and $m(2, \{2\}, 1, \{1\}) = R$. This is just the segregated equilibrium.